iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VII **Month of publication:** July 2022

DOI: https://doi.org/10.22214/ijraset.2022.44814

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

A Survey on Machine Learning Based Techniques for Sign Language Translation System

Harsh Shambhuwani[1], Ritik Dhabekar[3], Pranjali Gore[5], Mahima Tiwari[4], Bhushan Meshram[2], Milind Tote[6]
CSE/IT, JD College of Engineering & Management, Nagpur, India

Abstract: ***Sign language is the communication method used to bridge the gap among mute and deaf people and the rest of the people. But most people are not familiar with sign language, so communication becomes difficult among mute, deaf and the rest of the people. Improvement in machine learning technology has led to research on its application in sign language recognition and translation system. This paper presents a review of different sign language recognition and translation systems and the various techniques like classification algorithms that were implemented in them. We have found that although many sign language recognition and translation system exist none of them provides efficient and real-time conversation of sign language into audio and many sign language recognition and translation system are hand movement sensor-based which are not affordable to everyone. So, we have also discussed a sign language translation system integrated into a video calling application, in which we will be using a single shot multibox detector for hand detection, for feature extraction Inception V3 will be used, and support vector machines for classification, attention network for generating semantically correct sentences and a text to speech synthesizer for audio output. This system will be easy to access and affordable as it uses vision-based recognition, hence it will reduce the communication gap among mute, deaf and the rest of the people effectively.***
Keywords: sign language translation; video conferencing; text-to-speech

I. INTRODUCTION

In India, there are millions (approximately 63 million) of people who have hearing impairment according to an article of Hindustan times [1] and many others who are mute. These people face difficulty to communicate with rest of the people in the world on many occasions. To remove that communication gap among the normal people and the impaired person sign language is used. Sign language is a physical action which uses hands and eyes to communicate with deaf and mute people. They can express their feelings with different hand shapes and movements. The sign language recognition system acts as a communication medium among the hearing impaired and rest of the people. So real-time application is required to translate sign language to speech so that people will understand what a mute or a deaf person is trying to say/convey. And same application should also be applicable for an average person to communicate with a mute or a deaf person. The task is to convert their sign language into text and speech.

Many sign language recognition systems are there, but they do not give the proper real-time solution to convert sign language communication to voice or text, there are 2 types of method for sign language recognition where 1st method is hand movement sensor-based [2] (hand gloves to track hand movements) which are costly at the same time with lots of complexity and facial expressions can't be recognized & wearing a glove is most of the time uncomfortable for the users. As this type of method is expensive for the user it can't be implemented massively. Another method is vision-based recognition (use a camera), this method is easily accessible and relatively cheap as nowadays everyone carries at least one device which has a camera. The 2nd method is used in the vast number of papers in which the proposed applications/systems have different machine learning (ML) and deep learning (DL) approaches. In the later mentioned applications/systems the various algorithm used for classification purposes were support vector machines (SVM), Naive Bayers, logistic regression, artificial neural networks (ANN), convolution neural networks(CNN), k-nearest neighbours algorithm(KNN), etc. which gives different accuracies. The accuracy of correctly labelling sign gestures can be increased by using the most appropriate object detection and classification technique.

II. LITERATURE REVIEW

T RAGHUVEERA, et.al [4] presented a depth – based Indian Sign Language recognition using Microsoft Kinect. Authors used Speeded – Up Robust Feature (SURF) for segmentation and for feature extraction. The authors divided the work in five stages. First is pre – processing, then hand's segmentation, third is withdrawal of three features (SURF, oriented gradient's histogram and local binary patterns), then sign language recognition using SVM and at last prediction of output signs using ensemble technique & sentence interpretation.

©IJRASET: All Rights are Reserved

Aishwarya Sharma, et.al [6] presented Sign Language to Speech Translation. Authors proposed the real time application which translates the Sign Language into Speech Translation. Authors divided the work in 4 parts: In first section, capture the gestures as input for the interface, second - gesture recognition, third - translation of signs to text and then conversion of the text to speech and presenting an audio file as the output. In second section, authors had used SURF (Speed – Up Robust Features) and Scale Invariant Feature Transformation (SIFT) for feature extraction. In third section: CNN was used for classification; it helps to translate Sign Language to text. In last section, authors had used python libraries (Google Test-to-Speech) which helps to convert text to speech.

Malli Mahesh Chandra, et.al [8] presented Sign Language to Speech Conversion by applying the Support Vector Machines (SVM) classifier. with Authors are using Glove to take data from user. On these gloves flex sensors, gyroscope & accelerometer are embedded. Sensors capture the gestures of user's hands in real time. Arduino is used to collect the data from the sensors and forward it to the computer by using Bluetooth. The computer processes the data, classify the Sign language and predict the word with each gesture by using SVM Algorithm. This system is designed to predict both ISL and ASL gestures.

K. Revanth, et.al [10] presented Recognition of Sign language. Authors are using raw image of sign language to recognition of hand gestures. Completing the hand recognition, the skin masking process take place. The Feature Extraction is done by Oriented FAST and rotated Brief (ORB). Authors are using 25 set of commonly used signs datasets. When Feature Detection is completed, Clustering Process starts by using K – means. The received dataset which is clustered will be split into train and test files randomly. The percentage of splitting the dataset is fixed and decided as 60% - 40%. After Clustering, Classifier starts the process of classify by applying K – Nearest Neighbors (K-NN), SVM algorithm, Naive Bayes and Logistic Regression.

Tiantian Yuan, et.al [11] presented a sign language translation model inspired by sequence-to-sequence models and also introduced CSLD, a vast scale Chinese continuous sign language dataset. This dataset provides researchers with a large glossary of Chinese sign language (CSL) words in multiple condition. It contains multiple synchronized videos for different viewing angles, as well as annotations for the start and end times.

Muthu Mariappan H, et.al [13] presented a real time sign language recognition model by applying fuzzy c-means clustering machine learning algorithm for training & prediction purpose. Authors created their own data set for training & testing. OpenCV library was utilized for feature extraction and video classification. This FCM based sign language recognition system was correct 3/4th of the time in gesture labeling for recognizing the words of Indian sign language (ISL) and also requires more computation time.

Leela Surya Teja Mangamuri, et.al [15] presented that for sign language recognition segmentation of hand in this Two Hand Indian Sign Language plays a vital role in preprocessing. Skin filtering methods are proposed for extraction of hands from background. Skin colored pixels are segregated from non – skin-colored pixels by setting up a threshold. Adaptive Boost algorithm used for extracting two hands from the background.

Pengfei Sun, et.al [16] proposed an accurate and real time sign language recognition system. Using the key points of hands estimation method, they can acquire the hand data with precision. This method has strong robustness when working against some bad recognition conditions such as illumination changes or being sheltered by another objects. When compared with the popular recognition technique based on depth camera, this method has equal or even better accuracy, while the implementary difficulty can be greatly reduced. During training step, 5 kinds of machine learning algorithms are used for the classification of 20 letters in alphabet and the highest classification accuracy was realized by SVM and KNN algorithms.

M. I. N. P. Munasinghe [17] developed a vision-based system which can recognize hand gestures in front of a web camera in real time using motion history images (MHI) & feed forward neural networks. Implementations of this approach was done using python programming language and for image processing OpenCV (Open-Source Computer Vision Library) was used, for the machine learning tasks & to retain efficiency in large matrix manipulation tasks, python scikit-learn and NumPy libraries were used. In good lightning conditions the model achieved better performance as compared to performance bad lightning condition which shows that lightning condition plays significant role in accuracy of hand gestures recognition.

S. S Kumar, et.al [19] used a time series Recurrent Neural Network for ASL gloss recognition and a attention based sequence for translation of the recognition glosses into natural speech language. Gaussian Blur filter and Otsu's Binarization for segmentation and extraction of hands and face from the frame. Neural machine translation for natural language translation. The performance metrics- Gloss Error Rate (GER) & Gloss Recognition Rate (GRR) were used for evaluation purpose; gloss error rate was quite high (nearly ¼) due to these methods.

S Reshna, et.al [26] proposes an automatic Gesture recognition approach for Indian sign language (ISL) recognition. Skin color segmentation using YCbCr Skin color model to segment the hand region, feature extraction and Multi – class nonlinear Support Vector Machine proposes.

Authors had created the system to work in complex backgrounds by segregating the face and hand with respect to skin colour. The features which can identify the signs are withdrawn from the hand images and classified to recognize the sign. The classification is completed by using SVM algorithm. This work is successfully completed in python background.

Mohaned Hassan, et.al. [29] Presented a sign language recognition (SLR) system based on Hidden Markov Models (HMM) and a modified version of K- Nearest Neighbour (KNN). Authors had used two different datasets for their work. The first one is DG5-VHand data gloves. These gloves come with 5 bend sensors, one for each finger and the second one is Polhemus G4 motion sensors. Gloves & motion trackers are the most common sensors used for sensor-based SLR. Most of the time, measurements of these sensors are accurate enough thus feature extraction methods are not requisite.

A. Metrics

B.

Sr. No.	Author, year, & Publication	Techniques	Database Input	Performance
1.	Aishwarya Sharma, Dr. Siba Panda, Prof. Saurav Verma 2020 (IEEE)	Deep Neural Networks, Convolutional Neural Networks (CNN)	Data taken 120 gestures if the ISL taken. Data stored in image format	The range of accuracy was between 85-95%.
2.	T RAGHUVEERA, R DEEPTHI, R MANGALASHRI and R AKSHAYA 2020 (Springer)	Support Vector Machine (SVM), Speed up robust feature (SURF),Histogram of Oriented Gradient (HOG),Local binary Pattern(LBP)	The dataset used consists of 140 unique gestures of the ISL taken from 21 subjects, totalling to 4600 images.	The average recognition accuracy is up to 71.85% with this method.
3.	Leela Surya Teja Mangamuri, Lakshay Jain 2019 (IEEE)	Convolutional Neural Networks(CNN)	In it max pooling was used to get local maximum value of feature maps. It was done in 3 dimensions to accommodate video data.	This model achieved the test accuracy of 95.68% with 4.13% as false positive rate.
4.	Sujay S Kumar; Tenzin Wangyal; Varun Saboo 2019 (IEEE)	Time series neural networks	Dataset were obtained from the National center for Sign Language & Gesture Resources (NCSLGR) Corpus.	Recognition of individual glosses is 86% GRR with 23% GER
5.	Muthu Mariappan H & Dr.Gomathi V 2019 (IEEE)	Fuzzy c-means clustering ML algorithm	Dataset input was 80 words & 50 sentences of everyday usage terms of ISL by 10 different signers.	It has shown 75% accuracy in gesture labeling
6.	K. Revanth, N. Shri Madhava Raja 2019 (IEEE)	K- Nearest Neighbors	25 set of commonly used signs as image format	Accuracy 87.3742%.
7.	Malli Mahesh Chandra, Rajkumar S, Lakshmi Sutha Kumar 2019 (IEEE)	Support Vector Machine(SVM)	The received data corresponding to each gesture is continuously recorded at different positions. These recorded samples of each gesture are stored as a text file.	Accuracy of ISL 99% and ASL 98.91%.
8.	M. I. N. P. Munasinghe, 2018 (IEEE)	Feed forward neural network	Motion History Images were created as dataset by the author.	In good lightning- 85% success rate, in bad lightning- 71.3% success rate
9.	Pengfei Sun, Feng Chen, Guijin Wang, Jinsheng Ren, Jianwu Dong (2018) (Springer)	Machine Learning, CNN, KNN	Authors have collected 20 kinds of sign gestures, and each one has about 400–500 image samples.	The average recognition accuracy of 20 kinds of sign language is 95%
10.	S Reshna, M. Jayaraju 2017 (IEEE)	Support Vector Machine (SVM) and YCbCr color space	Dataset was created of 500 images (size: 200*200) for 11 symbols each.	The accuracy rate of this model is 85% to 90%.
11.	Mohaned Hassan, Khaled Assaleh, Tamer Shanableh 2016 (IEEE)	Hidden Markov Model (HMM) and K – Nearest Neighbour (KNN)	The Dataset input was 40 Arabic sentences with 80 words.	This model achieved 97% of accuracy rate in sentence recognition rate.
12.	Tiantian Yuan, Anesh Bhat 2016 (IEEE)	Recurrent neural network	The dataset recoded in video and stored in video format	Accuracy is up to 70-80%

Table 1

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

III. FUTURE WORK

In this section, we present a few ideas for future:

A video conference/calling app which will contain a sign language into audio translation feature which the user will be able to turn on/off by using a switch given on the user interface (UI). When this feature is turned on the live video will be fed into the sign language translator module in which SSD (single short multibox detector) is utilized for object detection, Inception v3 is applied for feature extraction and support vector machines(SVM) is utilized for classification & prediction purposes of the glosses by comparing with the dataset provided and then generate a text file of the identified glosses, when a word is associated with a sign it is called a gloss or we can say a gloss is a label. The set of the identified glosses will be converted into audio with the help of text-to-speech (TTS) synthesizer. A text-to-speech (TTS) synthesizer converts normal language text into artificial speech. The translated audio will be transmitted along with corresponding video from sign language user to other users. The ML model will be converted into a tflite (Tensorflow lite) model and integrated with the android application.

In the Web application we are going to use firebase for authentications, sign-in, sign-up, database (which contain user's user name, email id and password). Real-time database helps to store and synchronize data, fast and secured web hosting.

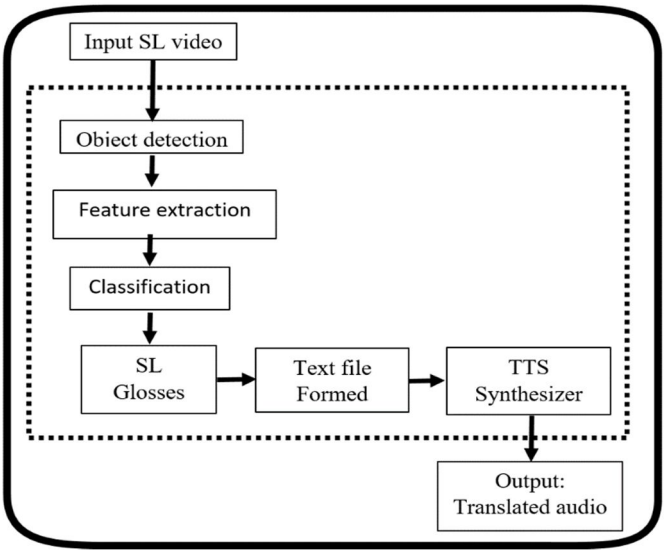


Fig 1. System Architecture of sign language translator

A. Object Detection

Single shot multibox detector (SSD) is an object detection algorithm, in this application it is used for hand detection. There are two models of SSD – SSD300 & SSD512. Here, SSD300 is used as it gives fast performance.

B. Feature Extraction

Here, Inception V3 is used for feature extraction, it was developed by a team at google. It is a convolutional neural network (CNN) for assisting in image analysis and object detection. It has a total of 42 layers & has a lower error rate than its predecessors (V1 & V2).

C. Classification

Support vector machines (SVM) is used for classification, SVM is one of the popular supervised learning algorithms. The fully connected layers of inception V3 is replaced with an SVM classifier.

IV. FLOW CHART

When the user starts the app, the Signup/Sign-in page will open on which the user will verify themselves or create a new account. After the verification user will land on home page having an option- start new meeting. The call will be connected when the callee accepts caller's call or click on the shared link. When the call is connected the users will land on conference room, the audio & video (allowed ones only) will be continuously transmitted & received by every user. At the bottom of the screen there will be option to turn on/off the sign language into audio translation feature.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VII July 2022- Available at www.ijraset.com

When this feature is turned on the live video will be fed into the sign language translator module in which single short multibox detector (SSD) will perform hand detection, Inception v3 will perform feature extraction on the detected hand gestures and support vector machines (SVM) will do classification & prediction of the glosses by comparing with the dataset provided and then a text file of the set of the identified glosses will be generated. The text fill will be converted in to audio. The translated audio will be transmitted along with corresponding video from sign language user to other users. The call will end when the user clicks on the end call option.

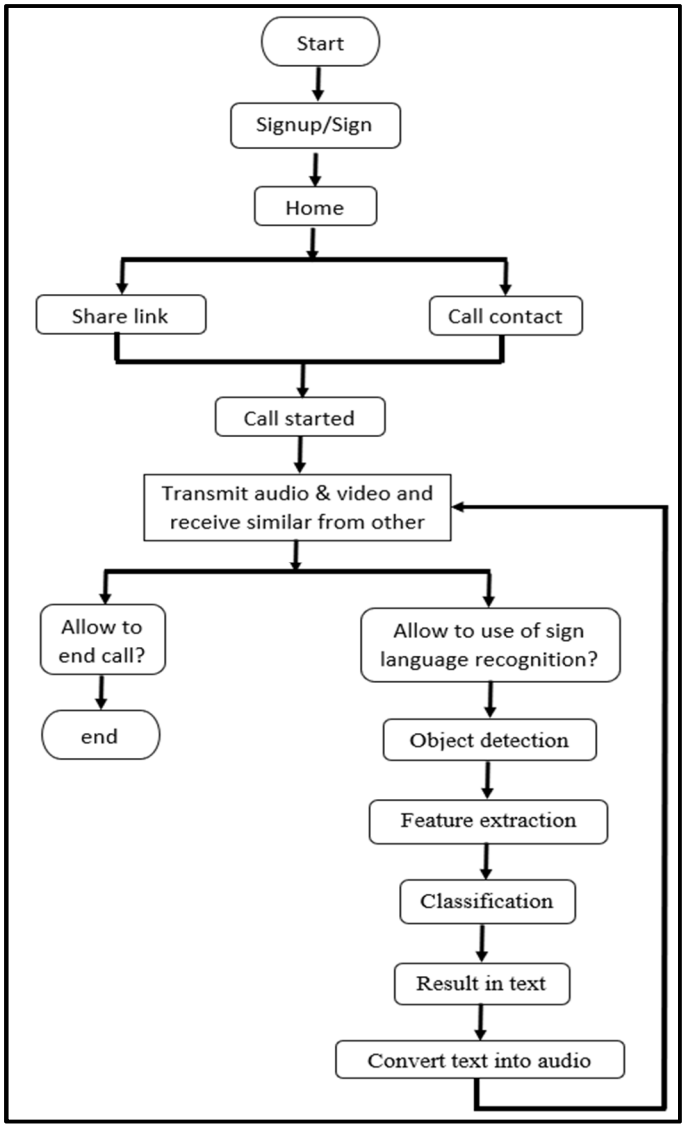


Fig. 2. Flow chart of the Application.

V. CONCLUSION

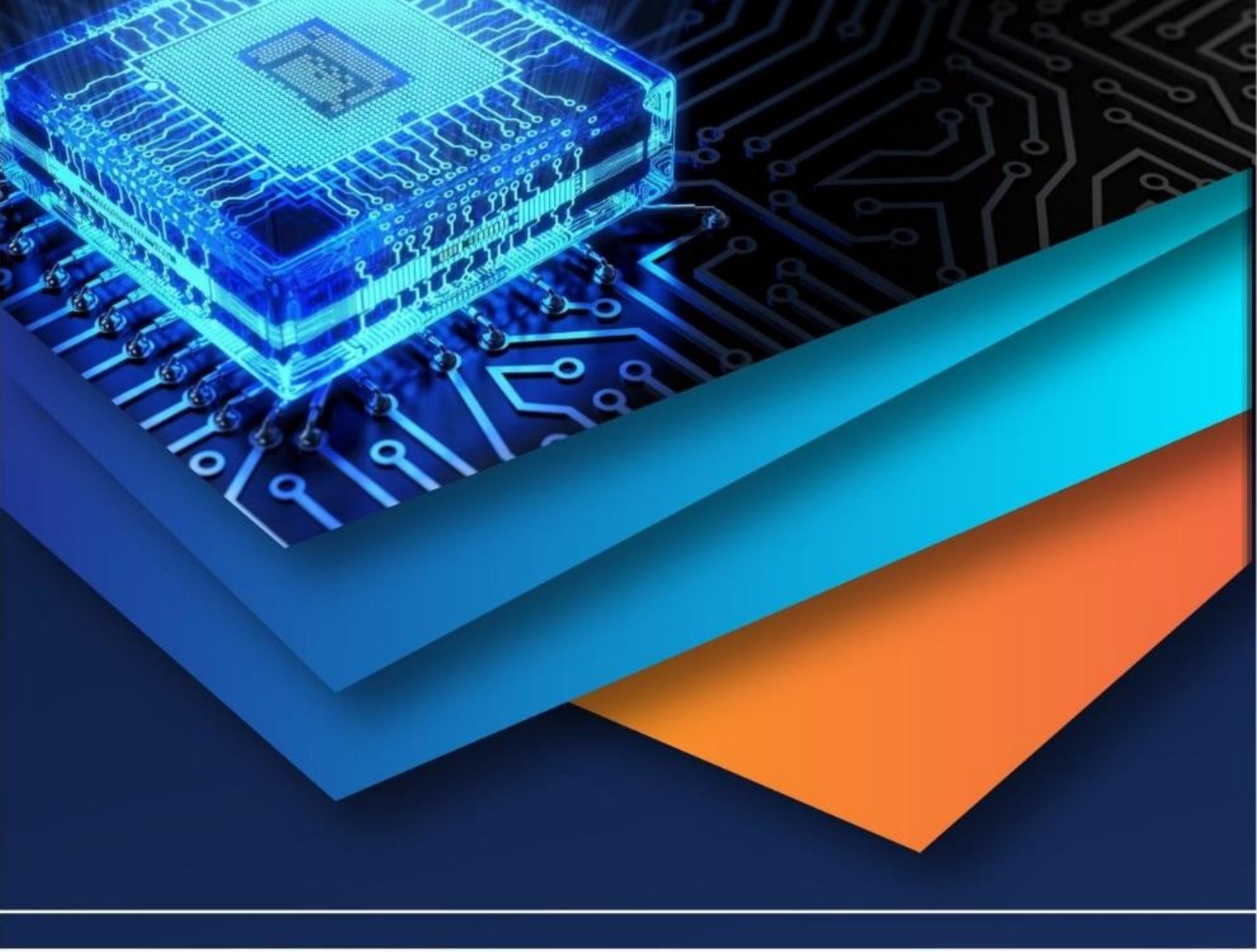
This paper presented a review of the sign language recognition system, which used machine learning algorithms and deep learning-based approaches. In the pursuit of a proper communication medium among hearing impaired /mute people, several gesture Recognition systems have been observed. Apart from Indian sign language (ISL), other language works are discussed with the classification algorithms and challenges. The future scope was also detected. Real-time SLR systems are still found to be having accuracy and fast translation problems. We have found out that the average accuracy of the reviewed systems is around 87%, so we will be using support vector machines (SVM) to improve the accuracy of the proposed model up to 95%(approximately). Our work will not only translate the sign language but will also convert it into audio so that another user on the receiver end can properly communicate with a mute person.

©IJRASET: All Rights are Reserved

REFERENCES

[1] https://www.hindustantimes.com/analysis/breach-the-wall-of-silence-give-state-recognition-to-indian-sign-language/story-hg7lj7LTWzfKgYB19prOhP.html
[2] N. Siddiqui and R. H. M. Chan, "Hand Gesture Recognition Using Multiple Acoustic Measurements at Wrist," in IEEE Transactions on Human-Machine Systems, vol. 51, no. 1, pp. 56-62, Feb. 2021.
[3] Razieh Rastgoo, Kourosh Kiani, Sergio Escalera " Sign Language Recognition: A Deep Survey" ELSEVIER, 2021
[4] T RAGHUVEERA, R DEEPTHI, R MANGALASHRI and R AKSHAYA," A depth-based Indian Sign Language recognition using Microsoft Kinect" Springer – 2020.
[5] Nimisha K P, Agnes jacob, "A Brief Of the Recent trends in Sign Language Recognition", International Conference on Communication and Signal Processing (ICCSP 2020)
[6] Aishwarya Sharma, Dr. Siba Panda, Prof. Saurav Verma,"Sign Language to Speech Translation",11th International Conference on Computing, Communication and Networking Technologies (ICCCNT 2020)
[7] P. Prajapati, Surya, G. Sai, and M. Nithya, "An Interpreter for the Differently Abled using Haptic Feedback and Machine Learning", in Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020).
[8] Malli Mahesh Chandra, Rajkumar S, Lakshmi Sutha Kumar, "Sign Languages to Speech Conversion Prototype using the SVM Classifier", IEEE Region 10 Conference (TENCON 2019)
[9] Palani Thanaraj Krishnan, Parvathavarthini Balasubramanian, "Detection of Alphabets for Machine Translation of Sign Language using Deep Neural Network", International Conference on Data Science and Communication (IconDSC 2019)
[10] K. Revanth, N. Shri Madhava Raja," Comprehensive SVM based Indian Sign Language Recognition", Proceeding of International Conference on Systems Computation Automation and Networking 2019
[11] Tiantian Yuan, Anesh Bhat, "Large Scale Sign Language Interpretation",14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)
[12] Surejya Suresh, Mithun Haridas T.P., "Sign Language Recognition System Using Deep Neural Network", 5th International Conference on Advanced Computing & Communication Systems (ICACCS 2019)
[13] Muthu Mariappan H & Dr. Gomathi V, "Real-Time Recognition of Indian Sign Language", Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019).
[14] Sandip Roy, Aarpan Kumar Maiti, Indira Gosh, Indranil Chatterjee, Kuntal Ghosh," A new Assistive Technology in Android platform to aid Vocabulary knowledge acquirement in Indian sign language for better reading comprehension in L2 and Mathematical ability", 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019.
[15] Leela Surya Teja Mangamuri, Lakshay Jain, Abhishek Sharmay, "Two Hand Indian Sign Language dataset for benchmarking classification models of Machine Learning",2th International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT 2019).
[16] Pengfei Sun, Feng Chen, Guijin Wang, Jinsheng Ren, Jianwu Dong, "A Robust Static Sign Language Recognition System Based on Hand Key Points Estimation", International Conference on Intelligent Systems Design and Applications, 2018.
[17] M. I. N. P. Munasinghe, "Dynamic Hand Gesture Recognition Using Computer Vision and Neural Networks", 3rd International Conference for Convergence in Technology (I2CT 2018).
[18] Shadman Shahriar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar Khan, Celia Shahmaz, Shaikh Anowarul Fattah, "Real Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning", (IEEE Region 10 Conference 2018).
[19] S. S Kumar, T. Wangyal, V. Saboo and R. Srinath, "Time Series Neural Networks for Real Time Sign Language Translation,"17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018.
[20] Mikel Labayen, Naiara Aginako, Basilio Sierra, Igor G. Olaizda, Julian Florez, "Machine Learning for video Action Recognition", 14th International conference on signal image technology and internet-based system (SITIS), 2018.
[21] Dasari Vishal, H M Aishwarya, K Nishkala, B Toshitha Royan, T K Ramesh, "Sign Language to Speech Conversion Using Smart Band and Wearable Computer Technology", IEEE International Conference on Computational Intelligence and Computing Research(ICCICR 2017)
[22] Shashank Salian, Indu Dokare, Dhiren Serai, Aditya Suresh, Pranav Ganorkar, "Proposed System for Sign Language Recognition", International Conference on Computation of Power Energy, Information and Communication (ICCPEIC 2017).
[23] S Yarisha Heera, Madhuri K Murthy, Sravanti V S, Sanket Salvi, "Talking Hands – An Indian Sign Language to Speech Translating Gloves", International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2017).
[24] Naresh Kumar, PhD Research Scholar, Department of Mathematics, "Sign Language Recognition for Hearing Impaired People based on Hands Symbols Classification, International Conference on Computing, Communication and Automation (ICCCA 2017)
[25] Mahesh M, Arvind Jayaprakash, Geetha M, "Sign Language Translator for Mobile Platforms", International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017.
[26] S Reshna and M. Jayaraju, "Spotting and Recognition of Hand Gesture for Indian sign language Recognition system with skin Segmentation and SVM", IEEE WiSPNET conference, 2017.
[27] Rudy Hartanto, Annisa Kartikasari, "Android Based- Real Time Static Indonesian Sign Language Recognition System with Prototype", 8th International Conference on Information Technology and Electrical Engineering (ICITEE), 2016.
[28] Bhuvan M S, Vinay Rao D, Siddharth Jain, Ashwin T S, Ram Mohana Reddy Guddetti, Sutej Pramod Kulgod, "Detection and Analysis Model for Grammatical Facial Expression in Sign Language", IEEE Region 10 Symposium (TENSYMP 2016).
[29] Mohaned Hasan, Khaled Assaleh, Tamer Shanableh, "User- Dependent Sign Language Recognition Using Motion Detection", International conference on computational science and computational Intelligence, 2016.
[30] Tejas Dharamsi, Rituparna Jawahar, Kavi Mahesh1 and Gowri Srinivasa, "Stringing Subtitles in Sign Language", IEEE 8th International Conference on Technology for Education(2016)

[31] Aadarsh K Singh, Benzil P John, Venkata Subramanian S R, Sathish Kumar A, Binoy B Nair,"A Low-cost Wearable Indian Sign Language Interpretation System", International Conference on Robotics and Automation for Humanitarian Applications (RAHA 2016)
[32] Mr.Sagar P.More, Prof Abdul Sattar,"Hand Gesture Recognition System Using Image Processing", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT 2016)
[33] V. Rusňák et al., "CoUnSiL: A video conferencing environment for interpretation of sign language in higher education," 2016 15th International Conference on Information Technology Based Higher Education and Training (ITHET), 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)